

## DAFTAR ISI

|  |      |
|--|------|
| PENGESAHAN.....  | ii   |
| PERNYATAAN.....  | iii  |
| PRAKATA.....   | iv   |
| ARTI LAMBANG DAN SINGKATAN.....                                  | vi   |
| ABSTRACT.....  | vii  |
| INTISARI.....  | viii |
| DAFTAR ISI.....  | ix   |
| DAFTAR GAMBAR.....   | xi   |
| DAFTAR TABEL.....  | xiii |
| BAB I PENDAHULUAN.....   | 1    |
| 1.1 Latar Belakang.....  | 1    |
| 1.2 Perumusan Masalah.....                                       | 5    |
| 1.3 Keaslian Penelitian.....                                     | 5    |
| 1.4 Tujuan Penelitian.....                                       | 10   |
| 1.5 Manfaat Penelitian.....                                      | 10   |
| 1.6 Batasan Penelitian.....                                      | 10   |
| BAB II TINJAUAN PUSTAKA DAN LANDASAN TEORI.....                  | 12   |
| 2.1 Tinjauan Pustaka.....  | 12   |
| 2.1.1 Analisis Pola Penumpang.....                               | 12   |
| 2.1.2 <i>Mining Smart Card Data</i> : Metode dan Pendekatan..... | 14   |
| 2.1.3 <i>Perbaikan dan Skalabilitas Metode Clustering</i> .....  | 16   |
| 2.2 Landasan Teori.....  | 18   |
| 2.2.1 <i>Data Warehouse</i> .....                                | 18   |
| 2.2.2 <i>Data Mining</i> .....                                   | 19   |
| 2.2.3 <i>Hadoop Platform</i> .....                               | 28   |
| 2.2.4 <i>Smart Card</i> .....                                    | 30   |
| 2.3 Pertanyaan Penelitian.....                                   | 32   |
| BAB III METODOLOGI.....  | 33   |
| 3.1 Alat dan Bahan.....  | 33   |
| 3.1.1 Alat.....  | 33   |
| 3.1.2 Bahan.....   | 33   |
| 3.2 Jalannya Penelitian.....                                     | 34   |
| 3.2.1 <i>Business Understanding</i> .....                        | 35   |
| 3.2.2 <i>Data Understanding</i> .....                            | 36   |
| 3.2.3 <i>Data Preparation</i> .....                              | 41   |
| 3.2.4 <i>Modelling</i> .....                                     | 50   |

|             |  |      |
|-------------|--|------|
| 3.2.5       | <i>Evaluation</i> .....  | 56   |
| 3.2.6       | <i>Deployment</i> .....  | 57   |
| 3.3         | Rancangan Penerapan <i>Preprocessing</i> Pada <i>Data Mining</i> .....             | 58   |
| BAB IV      | HASIL DAN PEMBAHASAN.....  | 59   |
| 4.1         | Hasil Penelitian.....  | 59   |
| 4.1.1       | Pengumpulan Data.....  | 59   |
| 4.1.2       | <i>Data Preprocessing</i> .....  | 59   |
| 4.1.3       | Penerapan <i>Scalable k-means++ Clustering</i> Dengan <i>Hadoop Platform</i> ..... | 63   |
| 4.2         | Evaluasi <i>Cluster</i> dan Pengujian.....   | 70   |
| 4.2.1       | <i>Sum of Squared Error (SSE)</i> .....  | 70   |
| 4.2.2       | <i>Silhouette Coefficient (SC)</i> .....   | 70   |
| 4.2.3       | Skenario Pengujian.....  | 72   |
| 4.2.4       | Hasil Pengujian.....   | 72   |
| 4.2.5       | Analisis Kendala dan Keterbatasan.....   | 77   |
| 4.2.6       | Konversi <i>File Output</i> Hasil <i>Clustering</i> .....                          | 78   |
| 4.3         | <i>Deployment</i> dan Visualisasi Hasil <i>Clustering</i> .....                    | 79   |
| BAB V       | KESIMPULAN DAN SARAN.....  | 85   |
| 5.1         | Kesimpulan.....  | 85   |
| 5.2         | Saran.....   | 86   |
| DAFTAR      | PUSTAKA.....   | 88   |
| LAMPIRAN 1  | .....  | L-1  |
| LAMPIRAN 2  | .....  | L-2  |
| LAMPIRAN 3  | .....  | L-3  |
| LAMPIRAN 4  | .....  | L-4  |
| LAMPIRAN 5  | .....  | L-5  |
| LAMPIRAN 6  | .....  | L-6  |
| LAMPIRAN 7  | .....  | L-7  |
| LAMPIRAN 8  | .....  | L-8  |
| LAMPIRAN 9  | .....  | L-9  |
| LAMPIRAN 10 | .....  | L-10 |
| IZIN        | PENELITIAN & PENGAMBILAN DATA.....   | L-11 |